

【质量与评价】

增值评价的基本理念、 实践困境与优化策略

李俊飞 谭顶良

【摘要】在教育公平与高质量发展理念的引导下,增值评价为教育评价提供了全新的视角和方法。增值评价理念扎根于教育公平与高质量发展观、以学评教的学生主体观和服务于改进的评价功能观,借助增值评价技术模型,从本质上探寻教育投入与产出的因果关系。但在实践中尚存一定的问题需要解决,主要包括技术模型尚不成熟、评价内容片面单一,以及评价结果反馈效果不佳等。未来可以通过系统测量和收集数据、妥善选用和改进增值评价模型,以及综合运用增值评价结果三个方面,使增值评价更为科学有效地开展,以更好地适应我国新时代教育评价改革和发展的需要。

【关键词】教育评价;增值评价;教育评价改革;增值模型

【作者简介】李俊飞,南京师范大学教育科学学院博士研究生;谭顶良,南京师范大学教育科学学院教授,博士生导师(江苏 南京 210097)。

【原文出处】摘自《中国教育旬刊》(京),2024.9.37~43

一、增值评价的理念诠释

(一)教育公平与高质量发展观

增值评价出现的缘由直指教育公平和高质量发展观。1966年詹姆斯·科尔曼向美国国会提交的《关于教育机会平等的报告》,利用量化手段指出了学校教育体系对学生实现成就上的作用日渐式微,呼吁将学校帮助学生学业进步与成长作为评价的标准,认为关注个体的发展和“投入—产出”的效益才是教育平等的根基。^[1]其将学生学业成就纳入教育机会公平的视野,直指纵向发展为核心的评价思路,间接催生了增值评价。此后,增值评价借助统计技术上的创新,落地成为一种新兴的教育评价方法,增值评价从概念的创生到实践应用,均以教育公平和高质量发展为目标。

促进教育公平、提高教育质量是我国当下对教育工作的总体要求。教育质量与公平二者已构成相互依存、相辅相成的关系。^[2]在差异不可避免的教育教学中,机会公平和过程公平同样重要,最终指向的是每位学生的潜力都得到充分的发展。^[3]可见,为每位学生提供适合的、全面发展的教育是全面提高教育质量与推进教育公平的基本内涵。教

育评价应当为其提供保障和动力。关注人的发展和为每个人提供适合的教育是实现更加公平、更高质量教育的重要路径。^[4]科学有效的教育评价需回归到人与教育的发展规律上,这与我国教育质量和公平的内涵具有高度的一致性。因此,教育评价不应该脱离“人”。另外,评价的关注点需要更多聚焦“个体”。教育评价不应该为了竞争,而更多的是为了合作,评价方法也应从粗放走向科学。评价内容应当充分体现学生的全面发展和个体差异,简单固化的应试已不再适应当下尊重主体差异的要求,也与教育高质量和公平观相违背。诸多评价内容忽略了教育的系统性和个体差异,导致评价不准确、不公平、不客观等问题,滋生了一种“目中无生”“分中无人”的片面实践。因此,我国的教育评价呼唤回归到“人”的主体性上,评价内容回归到“人”的全面发展上是应然之举,也是实现教育实质公平和高质量发展的必经之路。

增值评价作为一种较为新型的评价理念和方法,对于我国教育评价改革的“应然”目标有着足够的关照。增值评价对个体差异的关注尤为突出,其并非将固定的层次群体作为参照常模,而是使学生

自己跟自己比,更多关注“个体”的进步。通过纵向比较的形式,直指“以结果论英雄”的片面评价取向,改变利益相关者甚至整个社会对教育评价的关注点:从关注结果到关注过程、从关注条件到关注发展。^[5]当下,我国较为重视教育系统内的正义问题,而正义不仅是分配正义,还要保障每个社会成员享有独立、自由、平等的权利。^[6]增值评价则呼应了这种需求,通过关注过程、个体发展,以追求实质公平和高质量发展。再者,增值评价的一个基本前提是:真正优秀的学校或教师是对学生产生积极影响的,并促进他们的全面发展。增值评价的内涵也是在此前提下体现了公平比较的特征,即社会有权期望教师和学校为学生提供高质量和公平环境下的学习机会,无论学生进入任何一种教育场所,所有的学生都能够通过教育获得成功,并以每位学生在各阶段取得的进步来评判教师和学校的优劣。因此,增值评价的探索是我国教育评价改革的需要,也是通过评价作为“指挥棒”提升教育质量与公平的落脚点之一。

(二)“以学评教”的学生主体观

过去一段时间里,我国教育教学评价主要以“课堂中心”“教师中心”“书本中心”等理念为指导,通过比较目标与结果之间的差距,来评价教育质量。这种“以教评教”的方式在操作上较为扁平化,忽略了“教—学—评一致性”原则和学生主体发展的根本需要。随着学生主体地位的不断提升,“以学评教”的理念被提出,指以学生的学习效果来评价教师或学校的教育成效,强调以学生在学习中呈现的状况作为参照来评价教学的质量。^[7]

教育教学的目标常基于学生的表现而制订,教育教学对学生学习的影响程度也一直被认为是衡量有效教学的重要指标之一。^[8]之所以强调学生在教育评价中的地位,部分原因是人们会认为教师教只是影响学生学习的重要因素之一,而并非学生完成学习活动的必要条件,学生对教学的参与才是影响教育成效的关键。于是,“以学评教”中重要的指标是教与学的适当匹配。真正有效的教育教学应关注到学生学习、教师教学、环境支持等多方面的互动和联系。因此,不管是“以学评教”还是“以教评教”,都是在教与学一体的关照下以不同视角探究“什么是有效教学”的问题,这是“以学评教”的核心关切。

那么,回答增值评价为何根植于“以学评教”理念的问题便有据可循。首先,实现增值评价需要将

学生发展作为基础指标测算教师或学校效能,通过学生的阶段性进步,来确定教育教学的成效,这是根本前提。其体现了“以学评教”的基本内涵:以学生表现对教育教学实施评价。其次,增值评价的基本假设是教育教学对学生学习具有重要的影响,有效的评价应能够评估到学生主动参与学习的程度,同时在这个过程中也可以明晰教育教学对学生发展的贡献。借助统计模型,剥离其他非教育教学变量的干扰,更加精准地评估教育教学是如何影响学生学习的,并且对未来教与学的改进提供参考。最后,以学评教的理论体系尤其强调教学是否对不同水平的学生产生了显著的影响,这就需要将学生个体差异纳入考量。增值评价考虑到了学生个体差异对教学效能的影响,并使教、学、评三者可以在统计层面具有一致性。一对多的教学是目前较为常见的状态,增值模型在将学生个体差异变量纳入后,可以较为科学地测算每位学生在接受同一种教学后获得的增值,以回答“不同的学生在相同的教学下学习效果是否有差异”的问题。同时,教师或学校增值评价的基本逻辑就是捕捉教育系统的嵌套结构,即学生、班级、学校、学区等单元之间的层级性。增值评价不单单可以利用群体学生的平均增值量表示个体教师的教学效果,还可以判断在班级层面上教师的教学对不同学生的学业增值效应,以判断某个教学活动是否适应了学习者的个体差异。这种基于“以学评教”的增值评价理念通过比较学生纵向发展变化,来评价教师、学校等上维单元,体现了以学生为主体的教育评价观,同时兼顾了教育系统的嵌套结构,该结构以学生发展为中心向外拓展。

(三)服务于改进的评价功能观

增值评价的旨意可以追溯到第二代和第三代教育评价观。20世纪30年代,由拉尔夫·泰勒(Ralph W. Tyler)主导的“八年研究(1933-1940年)”中,正式提出了“教育评价”的概念,即“教育评价即是发现预设目标与实际结果之间的差距”,形成了以描述为主导的第二代评价观。^[9]随后,以判断为导向的第三代教育评价观应运而生,形成了“描述—判断”的教育评价模式。自第二代评价观提出以目标为核心的评价理念后,参照和比较成为评价的重要旨归,在现代教育评价理论中,明确了目标和结果之间的联结是评价得以执行的基本前提。同时,评价被认为是教育改进的一种工具,它

提供的信息可以让教育工作者确定哪些做法会实现预期的结果,哪些不会。总的来说,第二代、第三代教育评价观为增值评价确立了两大基准:一是比较构成评价,二是评价用于改进。

从增值评价的内涵追溯可以发现,比较是行为操作,改进是功能价值。传统理解中,教学与评价是两个相对独立的概念范畴。随着二者在理念和实践中的靠拢,教、学、评三者的关系逐步趋于紧密并达到一致。传统的评价功能观上,评价的“工具主义”高于“功能主义”,评价更重视问责和分流而非改进和激发,以“结果论英雄”和以“成绩论成败”的思想观念一直阻碍着素质教育的发展,使结果的达标取代了教育的根本追求——育人。此外,评价往往为了追寻形式公平,忽视了学生个体差异的客观事实。以至于评价陷入一种“部分人的成功是建立在其他人的失败之上”的畸形状态,很难关注到所有学生和教师甚至是学校的个体发展,埋没了师生的信心、成就和进步。因此,新的评价功能观要求从传统的结果问责转向关注全体学生的发展。

从功能观的转向可以看到增值评价在我国的探索并非“空穴来风”。增值评价的提出直指传统评价的主要弊端,即评价服务于高风险考核,而非改进和激发。增值评价兼具形成性评价和发展性评价的特征,既注重个体发展的过程,也关注对教育教学的持久支持。具体来说,可以通过分析不同学生、教师、学校的增值情况,发现提高教育教学质量的关键点。同时,结合结果评价中的指标,探究增值评价和结果评价如何关联,总结不同群体的优势和成因,实现因材施教和有针对性的教学改进。总之,与其说增值评价的探索是为了推动一种新的评价理念和技术,不如说是为了破除“结果至上、封闭竞争”的旧评价观提供新视角和新动力,引导教育评价功能观的转变,使社会关注每位教育主体的发展。因此,增值评价符合我国当下评价功能观转变的需要。

二、增值评价的实践困境

(一) 技术模型尚不成熟

增值评价需要借助增值模型较为客观地评价学生、教师和学校。增值评价的技术模型本质上是几种统计方法的统称,通过纵向追踪探寻教育投入与产出的过程,并推断增值中的因果律。模型的实现思路主要有两个层面。一是利用学生多年的学业成绩来纵向追踪其学业增值,收集学生在一段时

间内不同时间点上的测验结果,并对前后测成绩加以比较。二是关注学生的预期结果和实际结果之间的差距,测算教育教学使学生达到的实际进步是否高于学生在起点水平上可能达到的进步。^[10]两个层面相辅相成,互为支撑,考察“背景—投入—过程—产出”的过程,洞悉每位学生的进步程度。^[11]自增值评价出现以来,学界对增值评价技术模型的质疑从未停止,其中的重点之一就是模型自身的技术局限和应用模型的方法误区,这些问题给科学有效地开展增值评价实践带来了空前挑战。总体而言,增值评价模型的问题主要可以归纳为两个方面:一是统计假设,二是信度与效度。

模型的统计假设成立是科学有效应用技术模型的前提条件。最早有关增值评价模型的研究出现在农业、经济领域,这些领域应用增值模型的环境可以符合基本的统计假设,并且完全达到了数据的客观性要求,对增值的解释也相对直观可靠。^[12]但当模型进入教育教学领域则不再完全符合相应的前提假设,诸多人为主观的因素影响着增值模型的有效应用,这也是大部分研究者质疑增值评价模型科学性的原因之一。比如诸多增值的“值”并非单纯的线性数据,而呈现一种非线性的特征,违反线性假设可能导致增值结果不够精确。

增值评价面临的另一问题是评价的信度和效度问题。其一,信度主要与模型选择有关。目前常见的增值评价模型有线性回归模型、多水平回归模型、成长百分比模型等。面对诸多模型变式,不同的增值模型需匹配不同的应用条件,得到的增值结果也有差异。这种情况下,评价的利益相关者将会产生怀疑态度,势必会影响到增值评价的信度。其二,效度主要与协变量选择与模型精度有关。教育教学是一项十分复杂的活动,基于回归方法的增值模型需要对变量进行筛选。所有可能纳入的变量是否有合格的工具进行测量,是否还存在目前研究者无法进行观测但与教育质量紧密相关的变量,这些问题很难有定论。因此,选择协变量的主要困难在于无法捕捉和控制对教育教学效能有影响的所有变量。此外,当进行标准化测试时,测量误差自然会发生。测量可靠性的元分析结果表明,标准偏差可以达到0.09。^[13]这意味着教育测量中几乎每一个变量都存在一定程度的测量误差,也包括学业水平测试的分数。如果出现测不准的情况,也将影响回归模型的精度。

(二) 评价内容片面单一

首先,在教育系统中,学生、教师、学校、学区等模块的嵌套结构决定了学生是增值评价的核心角色。学生增值评价不仅要重视学生知识水平的增长,还要关注学生德、智、体、美、劳的全面发展。由于受增值评价模型统计假设上的限制,目前对学生的增值评价内容还是以测试分数为主,这可能忽略了学生综合素质发展,使教师和学校在追求高增值的过程中更加关注应试,进而加剧“唯分数”的导向,这是增值评价应用于学生评价的一个误区。^[14]增值评价的初衷在于破除“唯分数”的不良导向,如果评价内容固化为学生个体成绩的进步,虽然改变了横向比较的封闭状态,但也会落入另一种攀比成绩增值的恶性竞争处境中。

其次,利用学生成就评价教师也是增值评价的重要内容之一。在实践层面,教师增值评价更多关注教师的专业化发展和个体历时进程。所纳入的部分指标较为稳定,如职称、教龄、学历等,体现教师个体水平的发展规律,多属于课程结束后的外部指标,主要服务于教师的高风险绩效考核。然而,这种扁平化的内容指标忽略了诸多重要的教学因素,如教师的教学风格对学生发展的影响,也忽视了教师教学效能的评价内容。教师教学效能的高低取决于教师教学活动对学生发展的影响,根植于教和学的活动过程中,关注课程内部的过程性指标,探寻教学对学生全面发展的贡献,并反映教师的专业化发展和个体职业进程,体现教师个体水平的发展规律。^[15]受限于教学数据收集的困难,诸多较为复杂、灵活的教学效能指标难以纳入教师增值评价的内容体系中。这也使评价难以考察学生对教学的参与程度、动机和主观感受等潜在因素对教学质量的影响,同时也弱化了利用增值评价结果改进教学的功能。

最后,增值评价在对学校办学质量评价中的应用较为广泛。在评价学校办学时,增值评价的内容需要根据学校特征分门别类地进行选择。这就要求学校增值评价的内容不应仅反映逐年的升学率增值上,对学校办学质量的增值评价更需要扎根于学生全面发展和教师教学效能两个方面。由于学校处在教育嵌套体系的上维,学校办学质量需要体现出明显的学生中心观,在使用复杂增值模型进行计算时,学校全体学生和教师的全方位增值是办学效果的关键判断点。也就是说,利用增值评价模

型,将学校办学规模、投入资金等非教育教学变量的差异控制后,估计学校各方面办学条件对学生和教师发展所产生的影响。然而,当下学校增值评价受制于大数据体量和高复杂程度,评价内容难以深入办学的体制机制层面,致使校际的比较矮化为以学校为单位的升学增值。虽然可以一改既往的“强者恒强,弱者恒弱”的局面,有利于激励办学质量并不低却处在不利办学环境与条件下的学校,但这也反映出了“以结果论英雄”的终结性评价观念,阻碍了增值评价的科学开展。

(三) 结果反馈效果不佳

基于增值评价结果的反馈需要立足全过程、多层次、重诊断的原则。增值评价不同于传统的终结性评价,其过程耗时长、内容范畴广、结果形式多,容易出现低效甚至无效的评价反馈,从而对增值评价实践的整体效果产生不利影响。^[16]增值评价结果反馈效果不佳主要与增值结果不易理解以及忽视增值评价结果的诊断价值两个方面有关。

其一,理解部分增值评价结果是较为困难的。增值评价是一种借助各类统计模型进行纵向比较的评价方式,相比较传统标准化评价,其在评价理念和技术上具有一定的特殊性。一些增值评价的结果具有较为复杂、难以理解的问题,这类问题的出现与增值模型的统计方法和最终呈现的数据形式有关。如果师生没有相关的专业知识和技能,增长百分比、增值分数、效应量等多种具有复杂统计含义的结果将是难以理解的。在实践中,部分教师可能缺乏这方面的能力,导致结果解释不准确。增值评价虽可提供有用丰富的评价结果,但如果师生无法正确理解增值评价结果的含义,将会阻碍师生恰当地使用这些结果来改进教学或高风险决策。

其二,忽视运用增值评价结果来诊断改进教育教学。随着评价功能从证明转为改进,在促进教育教学的评价功能观下,对评价结果的应用决定了增值评价是否能有效发挥促进教育教学的作用。然而,师生和教育管理人员存在盲目应用增值评价结果来进行高风险绩效考核的问题。诚然,作为一种评价方式,高风险绩效考核是其一基本功能。而增值评价尚有一定的技术难题未得到妥善解决,增值评价结果的准确性尚待完善,应用增值评价结果作为高风险绩效考核的依据需要极为谨慎。从评价客体的角度看,接受评价者在高风险考核中如果并不信任评价结果,则会对增值评价结果产生怀疑

甚至抵触的心理,这就更加难以发挥促进教育教学的作用。因此,增值评价结果应当着眼于改进的目的而非绩效考核。不过应用评价结果开展教育教学改进的实践又是较为复杂的,师生难以理解通过统计模型得到的增值数据,亦会阻碍增值评价结果发挥改进和激发的作用,这就又回到了增值评价结果不易被接受的问题上。

三、应用增值评价的优化策略

(一)系统测量和收集数据

由于增值评价需要借助相应的统计模型来计算增值结果,就必须测量和收集一定规模的数据作为支撑。完备的数据支持是有效开展增值评价的基础,如果数据不够成熟,那么改善增值评价的策略便无从谈起。^[17]在收集数据之前,需要先全面系统地测量准确的数据。对于学业水平测试等标准化考试而言,这类数据是容易测得的。但如果涉及道德水平、审美素养等高维度变量时,测量会有一定的困难,传统的纸笔测验和自陈式量表已经无法完全代表学生的真实水平。因此,对于道德水平、审美素养、身心健康、劳动能力等方面的测量,需要更加关注学生多模态数据下的表现,并辅以表现性评价、质性评价等方面的结果作为参考。

此外,如果涉及教师或学校评价,也要考虑到学生、教师、学校等多个层面的数据。在收集数据的过程中,需要以学生、教师、学校三个层次分别规划变量集,如学生层的家庭经济社会水平、教师层的教龄、学校层的乡镇或城市区位等。这些都有可能影响最终增值结果的变量,需要分层处理,并从学生层面的结果层层推进至学校层面。通过数据分层可以搭建“学生发展—教师效能—学校办学”的评价证据链,以学生发展为中心,兼顾教师效能,在控制学校办学规模、投入资金等非教育教学变量的差异后,估计学校各方面办学条件对学生和教师发展所产生的影响。通过收集搭建系统的增值评价数据集,可以有效证明名校不一定适合所有学生,普通学校也不一定不适合所有学生。^[18]以评价教师教学效能为例,教学效能存在效果、效益和效率三个层面。^[19]数据集需要反映教师教学的投入和产出的关系。利用增值评价评估教师效能的实操重点是通过测量学生的发展,并结合系统的教师教学评价指标,估计教师对学生学习的影响。在应用过程中尽可能全面地收集教师的教学行为和个体特征变量,并利用学生的学业增值情况反映教师教

学效能,以对教师教学开展增值评价。

(二)妥善选用和改进增值评价模型

一方面,增值评价技术模型的选择是较多的,需要结合评价需要妥善选择。如果学生体量较大,则可以选择使用成长百分比模型,在区域范围内开展增值比较;如果学生体量较小并且需要开展相应的学习诊断,则使用可以纳入协变量的回归模型。在选择模型后需对不同评价内容作假设分析,部分高维变量可能并不符合所选增值模型的统计假设。比如道德、身心健康的发展不一定是线性增长的,其可能是有规律波动向上的非线性样态,这就需要结合学生评价的内容全面考虑选取适合的模型。对教师教学效能和学校办学的评价建议使用分层模型。不过分层模型的统计假设较为复杂,实践中能够在多大程度上满足前提假设,精准判别教师及其他影响因素对学生学习的影响,这是长期以来增值评价方法的争议所在。^[20]因此,在开展增值评价时需要结合评价项目的特征,综合评估教育教学活动有关的变量对学生学业发展的效果、效益和效率所产生的影响。这个过程中对一些高维度变量的测算需要使用符合数据特征的模型,各类模型的统计假设应得到充分检验。

另一方面,基于增值评价技术模型自限性问题,这里主要从减少统计假设的限制和提升模型的可解释性两个方面提出改进方向。通过减少统计假设的限制可以使模型更广泛地应用到不同的增值评价场景中,进而避免模型选择带来的不利影响。为此可以考虑进一步研究非线性模型和智能化模型在增值评价中的应用。相较于成绩作为因变量的传统线性模型来讲,核心素养等方面的增值建模或涉及非线性、潜变量测量、数据类别复杂等问题。解决这些问题的关键在于扩大模型的包容性和适应性,一种具有大规模并行处理、分布式信息储存、自组织、自学习等特点的智能化增值模型可能是未来增值评价研究的一个有希望的突破口。我国已有研究在理论上初步论证了这种智能化模型改进的适配性和工作流程。^[21]此外,提高模型的可解释性是指增进评价的利益相关者对模型内部运算机制与参数的理解,以提高测算结果的可信性。技术模型需要确保纳入运算的变量是有意义的。为了支撑不同的变量类型可以纳入模型中,需要对模型的入口结构做相应的调整,放宽对数据类型的限制,甚至减少标准化程序,以适应未来变量

多样化的特征。

(三)综合运用增值评价结果

针对增值评价结果难以理解和不够精准的问题,在此提出综合运用增值评价结果的优化策略。《深化新时代教育评价改革总体方案》中“四个评价”并不是相互独立的,结果评价、过程评价、增值评价、综合评价四者既是一个整体又相互联系,四者都是推动教育评价体系科学有效的关键路径。首先,增值评价的主要内容之一来自结果评价,也就是每一次学生在独立测试中获得的评价结果。从根本上讲,增值评价就是以纵向追踪的形式使用结果评价的信息。无论增值评价的内容如何,它都需要基于结果评价来判断增值,也就是增值比较的基准是由结果评价中得来。其次,过程评价是运用增值评价结果改进教育教学的载体,因为过程评价用于描述学生和教师的教与学过程,增值评价也是通过捕捉一段时间内的教育过程来开展诊断并为改进提供证据。最后,增值评价结果需要搭配其他评价方式的结果,也就是增值评价、结果评价以及过程评价等评价方式均要回归到综合评价的体系中。增值评价最好与其他教学评价结果结合使用,如来自课堂观察和专家评价的信息。这是使增值评价可以真正参与到“四个评价”的关键。因此,应将增值评价融入“四个评价”的体系中,不应以单一的增值评价结果作为主导。综合运用增值评价结果促使“四个评价”形成整体合力来推动我国教育评价改革,这或将是未来完善增值评价的重点之一。

最后需要强调的是定期开展对教师和教育管理者的专业技术培训是破解增值评价实践困境的关键。无论增值评价的数据如何完备、技术模型多么先进、评价内容是何,只有当教育工作者全面深入地了解增值评价的理念、技术、应用等方面的知识后,增值评价才能更为科学有效地落实,且真正成为一种可为的评价方式,以助力我国新时代教育评价的改革和发展。

参考文献:

- [1]王艳玲.教育公平与教师责任:《科尔曼报告》的启示——美国宾西法尼亚州立大学庞雪玲教授访谈[J].全球教育展望,2013,42(4):3-9.
- [2]袁振国.教育质量的国家观念[J].中国教育学报,2016(9):27-30.

[3]顾明远.个性化教育与人才培养模式创新[J].中国教育学报,2011(10):5-8.

[4]冯建军,高展.新时代的教育公平:政策路向与实践探索[J].东北师大学报(哲学社会科学版),2022(4):16-23.

[5]辛涛.“探索增值评价”的几个关键问题[J].中小学管理,2020(10):1.

[6]石中英.教育公平政策终极价值指向反思[J].探索与争鸣,2015(5):4-6.

[7]商发明.课堂教学评价新理念:以学评教[J].教育科学研究,2004(1):20-23.

[8]ZERI HUN Z, BEISHUIZEN J, OS W V. Student learning experience as indicator of teaching quality[J]. Educational Assessment Evaluation & Accountability, 2012, 24(2): 99-111.

[9]卢立涛.测量、描述、判断与建构:四代教育评价理论述评[J].教育测量与评价(理论版),2009(3):4-7.

[10]TEKWE C D, CARTER R L, MA C X, et al. An empirical comparison of statistical models for value-added assessment of school performance[J]. Journal of Educational and Behavioral Statistics, 2004, 29(1): 11-36.

[11]郑智勇,宋乃庆.新时代基础教育增值评价的三重逻辑[J].教育发展研究,2021,41(10):1-7,17.

[12]BRAUN H I. Using student progress to evaluate teachers: A primer on value-added models[M]. Princeton, N. J.: Educational Testing Service, 2005: 16.

[13]VACHA-HAASE T, THOMPSON B. Score reliability: A retrospective look back at 12 years of reliability generalization studies[J]. Measurement and Evaluation in Counseling and Development, 2011, 44(3): 159-168.

[14]李毅,郑鹏宇,黄怡铭.基础教育增值评价:内涵、价值与实证研究——以新时代中小学生阅读素养为例[J].中国电化教育,2022(10):47-55.

[15]孙颖.试析美国有效教师评价方式的价值取向:基于增值性评价和标准化评价[J].教育理论与实践,2015, 35(28):40-43.

[16]李俊飞,谭顶良.增值评价中的有效反馈问题探讨[J].中国考试,2023(5):29-36.

[17]李俊飞,谭顶良.增值评价模型应用的问题探讨与优化策略——以我国45项实证研究数据为佐证[J].中国考试,2024(6):71-78.

[18]任春荣.增值测量法:公平利用考试成绩评价学校效能的科学途径[J].中国考试(研究版),2007(4):12-16.

[19]姚利民.有效教学研究[D].上海:华东师范大学,2004.

[20]韩玉梅,严文蕃,蒋丹.探索增值评价的中国路向:基于美国实践经验的批判性分析[J].华东师范大学学报(教育科学版),2023,41(2):63-80.

[21]李俊飞,谭顶良,李格非.试析人工神经网络在增值评价中的应用[J].中国考试,2022(7):77-84.